**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**



Applicant: *VEGA, et al.*

Serial No.: 10/022,249

Filed: December 17, 2001

For: *HIGH THROUGHPUT DIRECTED EVOLUTION BY RATIONAL MUTAGENESIS*

Art Unit: 1643

Examiner: Unassigned

I hereby certify that this paper and the attached papers are being deposited with the United States Postal Service as first class mail in an envelope addressed to:
Commissioner for Patents
Washington, D.C. 20231, on this date.

03/01/2002
Date

Lynn Morkunas

**ATTACHMENT TO THE PRELIMINARY AMENDMENT**

**MARKED UP PARAGRAPHS AND CLAIMS**

**IN THE SPECIFICATION:**

Please amend the specification as follows:

**Please amend the paragraph on page 4, lines 6-17, as follows:**

The processes require accurate titering of the viruses in a collection or among collections (libraries) so that the activities of the screened mutant proteins can be compared. Provided are general methods for the quantitative assessment of the parameters of activity corresponding to the individual variants in the library, based upon intracellular serial dilution generated by precise [tittering]titering with the gene transfer viral [vectors]vectors. Any method that permits accurate titering may be used, including that described [nInternational]in International PCT application No. PCT/FR01/01366, based on French application n° 0005852, filed 9 May 2000, and published as International PCT application No. WO 01/186291. A method of titering, designated Tagged Replication and Expression Enhancement Technology (TREE™) is provided herein.

**Please amend the paragraph beginning on page 4, line 26, through page 5, line 19, as follows:**

A semi-rational method for evolution of proteins that is particularly designed for use in the methods herein or in any method that uses "evolved" proteins is also provided. The method, which is based on an amino-acid scanning protocol, is for rationally designing the variants for use in the directed evolution and selection method, and can employ iterative processing of the steps of the high throughput methods provided herein. In this method, once the target protein or domain is identified, nucleic acid molecules encoding variants are prepared. Each variant encoded by the nucleic acid molecules has a single amino acid replaced with another selected amino acid, such as alanine (Ala), glycine (Gly), serine (Ser) or any other suitable amino acid, typically one selected to have a neutral effect on secondary and tertiary structure. The resulting series of variants are separately screened in the high throughput format provided herein, and those that have a change in the target activity are selected and the modified amino acids are designated "hits." Nucleic acid molecules encoding proteins in which each hit position is replaced by the eighteen remaining amino acids then are synthesized and the resulting collection of molecules [are]is screened, such by introduction into host cells, and the proteins that result in [a]an improvement of a targeted activity, are identified. Such proteins are designated "leads." Leads may be further modified by producing proteins that have combinations of the mutations identified in the leads. This method, which does not require any knowledge of the structure of a target protein, permits precise control of locations where changes are introduced and also the amount of change that is introduced.

**Please amend the paragraph on page 6, lines 4-7, as follows:**

Protein/protein domain variants identified using the methods are also provided. Also provided are nucleic acid molecules and proteins and

polypeptides produced by the methods and viruses and cells that contain the nucleic acid [moleucles]molecules and proteins.

**Please amend the paragraph on page 7, lines 2-17, as follows:**

[FIGURES 1]FIGURES 1A - 1E summarize various exemplary embodiments of the high throughput processes provided herein. FIGURE 1A depicts an embodiment of the process in which an amino acid scan is employed to generate a library of mutants, which are then introduced into viral vectors, such as an adeno-associated viral vector (AAV), [an]a herpes virus, such as herpes simplex virus (HSV) and other herpes virus vectors, a vaccinia virus vector, retroviral vectors, such as MuMLV, MoMLV, feline leukemia virus, and HIV and other lentiviruses, adenovirus vectors and other suitable viral vector, each member of the library is individually tested and phenotypically characterized to identify HITS. FIGURE 1B summarizes round 2 in which LEADS are developed by mutagenesis at and/or surrounding the positions identified as HITS; FIGURE 1C summarizes the optional next round in which recombination among LEADS is performed to further optimize the LEADS; FIGURE 1D depicts the process in mammalian cells; and FIGURE 1E depicts the process in bacterial cells.

**Please amend the paragraphs beginning on page 8, line 24, through page 9, line 2, as follows:**

As used herein, directed evolution refers to [mehods]methods that [adapt"]adapt natural proteins or protein domains to work in new chemical or biological environments and/or to elicit new functions. It is more a more broad-based technology than DNA shuffling.

As used herein, high-throughput screening (HTS) refers to processes that test a large number of samples, such as samples of test proteins or cells containing nucleic acids encoding the proteins of interest to identify structures of interest or [the]to identify test compounds that interact with the variant

proteins or cells containing them. HTS operations are amenable to automation and are typically computerized to handle sample preparation, assay procedures and the subsequent processing of large volumes of data.

**Please amend the paragraph beginning on page 9, line 19, through page 10, line 3, as follows:**

As used herein, "leads" are "hits" whose activity has been optimized for the particular attribute, chemical, physical or biological property. In the methods herein, leads are generally produced by systematically replacing the hit loci with all remaining 18 amino acids, and identifying those among the resulting proteins that have a desired activity. The leads may be further optimized by replacement of a plurality of "hit" residues. Leads may be generated by other methods known to those of skill in the art and tested by the highthroughput methods herein. For purposes herein a lead typically has activity with respect to the function of interest that differs from the native activity, by a desired amount and [is at] by at least 10%, 20%, 30% or more from the wild type or native protein. Generally a Lead will have an activity that is 2 to 10 or more times the native protein for the activity of interest. As with hits, the change in the activity is dependent upon the activity that is "evolved." The desired alteration will depend upon the function or property of interest.

**Please amend the paragraph on page 11, lines 4-20, as follows:**

As used herein, adeno-associated virus (AAV) is a defective and non-pathogenic parvovirus that requires co-infection with either adenovirus or herpes virus for its growth and multiplication, able of providing helper functions. A variety of serotypes are known, and contemplated herein. Such serotypes include, but are not limited to: AAV-1 (Genbank accession no. NC002077; accession no. VR-645); AAV-2 (Genbank accession no. NC001401; accession no. VR-680); AAV-3 (Genbank accession no. NC001729;[acession]accession

no. VR-681); AAV-3b (Genbank accession no. NC001863); AAV-4 (Genbank accession no. NC001829; ATCC accession no. VR-646 ); AAV-6 (Genbank accession no.[NC001729]NC001862); and avian associated adeno-virus (ATCC accession no. VR-1449). The preparation and use of AAVs as vectors for gene expression *in vitro* and for *in vivo* use for gene therapy [is]are well known (see, *e.g.*, U.S. Patent Nos. 4,797,368, 5,139,941, 5,798,390 and 6,127,175; Tessier *et al.* (2001) *J. Virol. 75:*375-383; Salvetti *et al.* (1998) *Hum Gene Ther 20*:695-706; Chadeuf *et al.* (2000) *J Gene Med 2*:260-268).

**Please amend the paragraph beginning on page 12, line 29, through page 13, line 2, as follows:**

As used herein, $\epsilon$(efficiency), is the slope at the [inflexion]inflection point of the Hill curve (or, in general, of any other sigmoidal or linear approximation), to [asses]assess the efficiency of the global reaction (the biological agent and the assay system taken together) to elicit the biological or pharmacological response.

**Please amend the paragraph on page 13, lines 11-14, as follows:**

As used herein, a library of mutants refers to a collection of plasmids or other vehicles that [carrying (encoding)]carry (encode) the gene variants, such that individual [plasmid]plasmids or other vehicles carry individual gene variants. When a library of proteins is contemplated, it will be so-stated.

**Please amend the paragraph on page 13, lines 18-24, as follows:**

As used herein, "reporter" or "reporter moiety" refers to any moiety that allows for the detection of a molecule of interest, such as a protein expressed by a cell. Reporter moieties include, but are not limited to, for example, fluorescent proteins, such as red, blue and green fluorescent proteins; lacZ and other detectable proteins and gene products. For expression in cells, nucleic

acid encoding the reporter moiety can be expressed as a fusion protein with a protein of interest or under [to] the control of a promoter of interest.

**Please amend the paragraph on page 14, lines 1-7, as follows:**

As used herein, activity refers to the function or property to be [evolved]evolved. An active site refers to a site(s) responsible or that participates in conferring the activity or function. The activity or active site evolved (the function or property and the site conferring or participating in conferring the activity) may have nothing to do with natural activities of a protein. For example, it could be an 'active site' for conferring immunogenicity (immunogenic sites or epitopes) on a protein.

**Please amend the paragraphs beginning on page 16, line 36, through page 18, line 2, as follows:**

As used herein, nucleic acids include DNA, RNA and analogs thereof, including protein nucleic acids (PNA) and mixture thereof. Nucleic acids can be single or double stranded. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 14, 16, 20, 30, 50, 100 or more nucleic acid bases long.

As used herein, [by] homologous means about greater than 25% nucleic acid sequence identity, preferably 25% 40%, 60%, 80%, 90% or 95%. The intended percentage will be specified. The terms "homology" and "identity" are often used interchangeably. In general, sequences are aligned so that the highest order match is obtained (see, *e.g.*: *Computational Molecular Biology*,

Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; Carillo *et al.* (1988) *SIAM J Applied Math 48*:1073). By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate stringency or at high stringency all along the length of the nucleic acid of interest. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

As used herein, a nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a therapeutic polypeptide. By the term "substantially homologous" it is meant having at least 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

**Please amend the paragraph beginning on page 20, line 26, through page 21, line 17, as follows:**

As used herein, genetic therapy involves the transfer of heterologous nucleic acids to the certain cells, target cells, of a mammal, particularly a human, with a disorder or conditions for which such therapy is sought. The nucleic acid, such as DNA, is introduced into the selected target cells in a manner such that the heterologous nucleic acid, such as DNA, is expressed and

a therapeutic product encoded thereby is produced. Alternatively, the heterologous nucleic acid, such as DNA, may in some manner mediate expression of DNA that encodes the therapeutic product, or it may encode a product, such as a peptide or RNA that in some manner mediates, directly or indirectly, expression of a therapeutic product. Genetic therapy may also be used to deliver nucleic acid encoding a gene product that replaces a defective gene or supplements a gene product produced by the mammal or the cell in which it is introduced. The introduced nucleic acid may encode a therapeutic compound, such as a growth factor or inhibitor thereof, or a tumor necrosis factor or inhibitor thereof, such as a receptor therefor, that is not normally produced in the mammalian host or that is not produced in therapeutically effective amounts or at a therapeutically useful time. The heterologous nucleic acid, such as DNA, encoding the therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof. Genetic therapy may also involve delivery of an inhibitor or repressor or other modulator of gene expression.

**Please amend the paragraph on page 23, lines 1-8, as follows:**

Thus, by "isolated" it is meant that the nucleic is free of the coding sequences of those genes that, in the naturally-occurring genome of the organism (if any), immediately flank the gene encoding the nucleic acid of interest. Isolated DNA may be single-stranded or double-stranded, and may be genomic DNA, cDNA, recombinant hybrid DNA, or synthetic DNA. It may be identical to a native DNA sequence, or may differ from such sequence by the deletion, addition, or substitution of one or more nucleotides.

**Please amend the paragraph on page 24, lines 15-23, as follows:**

As used herein, the phrase "operatively linked" generally means the sequences or segments have been covalently joined into one piece of DNA,

whether in single or double stranded form, whereby control or regulatory sequences on one segment control or permit expression or replication or other such control of other segments. The two segments are not necessarily contiguous. For gene expression a DNA sequence and a regulatory sequence(s) are connected in such a way to control or permit gene expression when the appropriate [molecular]molecules, e.g., transcriptional activator proteins, are bound to the regulatory sequence(s).

**Please amend the paragraph on page 25, lines 10-23, as follows:**

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form, are not bound to the chromosome. "Plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. Other such [other] forms of expression vectors that serve equivalent functions and that become known in the art can be used subsequently hereto.

**Please amend the paragraph on page 29, lines 10-14, as follows:**

Provided herein are high throughput [process]processes for the generation of and identification of proteins that exhibit desired phenotypes. The [processes,]processes include methods that are particularly adapted for high throughput protocols, which require accurate methods for identifying modified proteins.

**Please amend the paragraphs on page 30, lin s 11-16, as follows:**

Generation of diversity at the nucleic acid level, in principle, can be accomplished by a number of diverse technologies like mutagenesis (either site-directed or random), recombination, shuffling and de-novo synthesis. These different technologies differ in the degree of diversity they generate as well as in the minimal length of the unitary change they can introduce (from single base to large domains). The outcome of step 1 is a collection of diverse, although highly related, molecules that constitutes what is called a 'library'.

**Please amend the paragraph beginning on page 33, line 27, through page 34, line 5, as follows:**

The method includes the following steps:

(1) Identification of the active site(s) on the full length protein sequence. In one embodiment a full-length amino acid-scan, typically, although not necessarily, an Ala-scan, or the identification and positioning of the active site(s) on proteins of either known or unknown function. For purposes herein, an active site is not necessarily the natural active [sites]site involved in the natural activity of a target protein, but those amino acids involved in the activities of the proteins under 'directed evolution' with the purpose of either gain, improvement or loss of function.

**Please amend the paragraph on page 36, lines 13-27, as follows:**

The technologies available for gene transfer and expression into either bacteria or eukaryotic (let's consider mammalian) cells widely vary in their intrinsic efficiencies. While it is very easy to efficiently transfer and express genes in bacteria by chemical/physical methods (transformation), that is not the case for mammalian cells, where the transformation (here called transfection) process is inefficient and unreliable, [specially]especially when reproducibility and robustness are necessary in miniature, large [number-]number and small

scale high throughput settings like those necessary to analyze gene variant libraries. Therefore, when transfection is used on mammalian cells, the specific activity measured for the individual variants in the library most probably does not accurately reflect the real specific activity of the molecules involved. As provided herein, transduction, the process of gene transfer and expression into mammalian and other cells mediated by viruses, overcomes the limitations of transfection.

**Please amend the paragraphs beginning on page 37, line 25, through page 38, line 2, as follows:**

Much progress in gene therapy, genomics, biotechnology and, in general, biomedical sciences, depends on the ability to generate and analyze large numbers and small amounts of specific viruses. High throughput technologies are employed in disciplines such as functional genomics and gene therapy in which the use of viruses plays a key role for the efficient transfer and analysis of large collections of genes or libraries. Also, virus samples and biomedical samples containing viruses are routinely analyzed in thousands of hospitals, health centers, academic labs and biotech [setting]settings.

**Please amend the paragraphs beginning on page 39, line 17, through page 41, line 4, as follows:**

Real Time Virus Titering [(RTVT)](RTVT™) published as International PCT application No. WO 01/186291, which is based on PCT/FR01/01366 and EXAMPLES below) uses non-destructive methods for the assessment of output signal. Real Time Virus Titering is a viral titration method based on the kinetic analysis of the development of the output signal in virus-infected cells, tested at a single concentration of virus or biological sample. Instead of fixing the time point after infection and varying the concentration of the sample as is done in limiting dilution methods, in the Real Time Virus [Titerng]Titering RTVT™ method,

a fixed concentration of virus is used and the generation of a signal over time is assessed. Hence the signal is measured as a function of time at a single virus concentration. In this situation, a single virus sample (concentration), whose output signal is measured at a number of time points, can give rise to as many measurements of the output signal as needed, and, eventually to a continuous, over time, assessment of the signal in real time.

Real Time Virus [Titerng]Titering RTVT™ can be advantageously used in high throughput methods in which large numbers of biological samples are analyzed at the same time. This is the case, for instance, when titering viruses in a virus library. Limiting dilution methods rely on the output signal generated by a number of dilutions of each individual sample. If, for example, 10 dilutions (or experimental points) of each virus are used for a titration using a limiting dilution method, the analysis of a library containing 10,000 viruses require analysis of $10^5$ (*i.e.,* 10 x 10,000) experimental points. The Real Time Virus [Titerng]Titering RTVT™ method requires only one dilution per sample, thereby requiring 10-fold fewer experimental points than a limiting dilution method. For a Real Time Virus [Titerng]Titering RTVT™ titering system, the time ($t\beta$) necessary for the output signal to reach a reference value ($\beta$) is a direct function of the concentration of virus. Thus, $t\beta$ can be used to directly determine the concentration of the virus.

A limitation of the Real Time Virus [Titerng]Titering RTVT™ limiting dilution titering method, however, is that not all the viruses (nor the genes carried by the viruses) generate a readily measured output signal that can be followed over time using non-destructive methods.

**Tagg d Replication and expression enhancement (TREE)**

A method for titering designated Tagged Replication and Expression Enhancement Technology (TREE™) is provided herein. This system includes: i) a cell, ii) a reporter virus carrying a reporter gene, whose activity can be followed over time by a non- destructive method (*i.e.*, fluorescence), iii) the virus (or virus library to be titered), herein referred to as the "titering virus". The elements are employed such that the virus to be titered interferes with any output signal generated by the reporter virus, leading to either decrease or increase in the amount of that signal. The higher the amount of virus to be titered, the higher is the interference with the reporter virus and output signal. In the absence of virus to be titered, the kinetics of the output signal generated by the reporter virus are followed using the Real Time Virus [Titerng]Titering RTVT™ titering method. In the presence of increasing amounts of the virus to be titered the output signal takes longer (or shorter) to develop as a function of the amount of virus to be titered.

**Please amend the paragraph beginning on page 43, line 29, through page 44, line 3, as follows:**

The mutated forms of the nucleic acid are prepared or introduced into plasmids for expression in bacterial cells. The genetic variants are expressed from suitable bacterial cells, which are prepared by transformation of aliquots of the cells with each member of the plasmid library (each genetic variant continues to be physically separated from each other).

**Please amend the paragraph on page 44, lines 12-19, as follows:**

(2) Titration of the virus library (of each individual virus present in the library, separately). Titration is effected by any method, but generally by either a method designated Real Time Virus Titering (RTVT™) (see, International PCT application No. PCT/FR01/01366 published as International PCT application No. WO 01/186291 and the EXAMPLES below) or a refinement

of that method [provide]provided herein and designated Tagged Replication and expression enhancement (TREE™) described above and in the examples;

**Please amend the paragraphs beginning on page 47, line 29, through page 49, line 7, as follows:**

Regulation of AAV genes is complex and involves positive and negative regulation of viral transcription. For example, the regulatory proteins Rep 78 and Rep 68 interact with viral promoters to establish a feedback loop (Beaton *et al.* (1989) *J. Virol 63*:4450-4454; Hermonat (1994) *Cancer Lett 81*:129-136). Expression from the p5 and p19 promoters is negatively regulated in *trans* by these proteins. Rep 78 and 68, which are required for this regulation, [have] bind to inverted terminal repeats (ITRs; Ashktorab *et al.* (1989) *J. Virol. 63*:3034-3039) in a site- and [stand-]strand-specific manner, *in vivo* and *in vitro*. This binding to ITRs induces a cleavage at the TRS and permits the replication of the hairpin structure, thus, illustrating the Rep helicase and endonuclease activities (Im *et al.* (1990) *Cell 61*:447-457; and Walker *et al.* (1997) *J. Virol. 71*:6996-7004), and the role of these non-structural proteins in the initial steps of DNA replication (Hermonat *et al.* (1984) *J. Virol. 52*:329-339). Rep 52 and 40, the two minor forms of the Rep proteins, do not bind to ITRs and are dispensable for viral DNA replication and site-specific integration (Im *et al.* (1992) *J. Virol. 66*:1119-112834; Ni *et al.* (1994) *J. Virol. 68*:1128-1138.

The genome (see, FIG. 4) is organized into two open reading frames (ORFs, designated left and right) that encode structural capsid proteins (Cap) and non-structural proteins (Rep). There are three promoters: p5 (from nucleotides 255 to 261: TATTTAA), p19 (from nucleotide 843 to 849: TATTTAA) and p40 (from nucleotides 1822 to 1827: ATATAA). The right-side ORF (see FIG. 4) encodes three capsid structural proteins (Vp 1-3). These three proteins, which are encoded by overlapping DNA, result from differential splicing and the use of an unusual initiator codon (Cassinoti *et al.* (1988) *Virology*

*167*:176-184). Expression of the capsid genes is regulated by the p40 promoter. Capsid proteins VP1, VP2 and VP3 [intiate]initiate from the p40 promoter. VP1 uses an alternate splice acceptor at nucleotide 2201; whereas VP2 and VP3 are derived from the same transcription unit, but VP2 use an ACG triplet as an initiation codon upstream from the start of VP3. On the left side of the genome, two promoters p5 and p19 direct expression of four regulatory proteins. The left flanking sequence also uses a differential splicing mechanism (Mendelson *et al.* (1986) *J. Virol 60*:823-832) to encode the Rep proteins, designated Rep 78, 68, 52 and 40 on the basis molecular weight. Rep 78 and 68 are translated from a transcript produced from the p5 promoter and are produced from the unspliced and spliced form, respectively, of the transcript. Rep 52 and 40 are the translation products of unspliced and spliced transcripts from the p19 promoter.

**Please amend the paragraph beginning on page 49, line 20, through page 50, line 5, as follows:**

AAV and rAAV have many applications, including use as a gene transfer vector, for introducing heterologous nucleic acid into cells and for genetic therapy. Advances in the production of high-titer rAAV stocks to the transition to human clinical trials have been made, but improvement of rAAV production will be complemented with special attention to clinical applications of rAAV vectors as a successful gene therapy approach. Productivity of rAAV (i.e. the amount of vector particles that can be obtained per unitary manufacturing peration) is one of the rate limiting steps in the further development of rAAV as a gene therapy vector. Methods for high throughput production and screening of rAAV have been developed (see, *e.g.*, Drittanti *et al.* (2000) *Gene Therapy* [7:942-929]7:924-929). Briefly, as with the other steps in methods provided herein, the plasmid preparation, transfection, virus productivity and titer and biological activity assessment are intended to be performed in an automatable high throughput format, such as in a 96 well or loci formats (or other number of

wells or multiples of 96, such as 384, 1536 . . . 9600, 9984 . . well or loci formats).

**Please amend the paragraph on page 50, lines 6-23, as follows:**

Since the Rep protein is involved in replication it can serve as a target for increasing viral production. Since it has a variety of functions and its role in replication is complex, it has heretofore been difficult to identify mutations that result in increase viral production. The methods herein, which rely on *in vivo* screening methods, permit optimization of its activities as assessed by increases in viral production. Provided herein are Rep proteins and viruses and viral vectors containing the mutated Rep proteins that provide such increase. The amino acid positions on the rep proteins that are relevant for rep proteins activities in terms of AAV or rAAV virus production are provided. Those amino acid [position]positions are such that a change in the amino acid leads to a change in protein activity either to lower activity or increase activity. As shown herein, the alanine or amino acid scan revealed the amino acid positions important for such activity (i.e. hits). Subsequent mutations produced by systematically replacing the amino acids at the hit positions with the remaining 18 amino acids produced so-called "leads" that have amino acid changes and result in higher virus production. In this particular example, the method used included the following specific steps.

**Please amend the paragraph on page 51, lines 4-8, as follows:**

In a second experimental round, which included a new set of mutations and phenotypic analysis, each amino acid position hit by the Ala-scan step, was mutated by amino acid replacement of the native amino acid by the remaining 18 amino acids, using [site directed-mutagenesis]site-directed mutagenesis.

**Please amend the paragraph on page 51, lines 13-21, as follows:**

A plasmid library was thus generated in which each plasmid contained a different mutant bearing a different amino acid at a different hit position. Again, each resulting mutant rep protein was then expressed and the amount of virus it

could [produced measure]produce measured as indicated below. The relative activity of each individual mutant compared to the native protein is indicated in FIGURE 3B. Leads are those mutants that lead to an increase in the activity of the protein (in the example: the ten mutants with activities higher, typically between 2 to 10 times or more, generally 6-10 time, than the native activity).

**Please amend the paragraph on page 55, lines 2-11, as follows:**

Toxicity and therapeutic efficacy of the rAAV can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the $LDS_{50}$ (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio $LD_{50}/ED_{50}$. Doses that exhibit large therapeutic indices are preferred. Doses that exhibit toxic side effects may be [used,]used; care should be taken to design a delivery system that targets rAAV to the site of treatment in order to minimize damage to untreated cells and reduce side effects.

**Please amend the paragraph on page 56, lines 22-31, as follows:**

**Serial dilution methods**

The assessment of the concentration or titer of biological agents using current approaches needs [for] serial dilutions of the agent. Serial dilutions of the agent are applied to a cell-based [reported]reporter system, that elicits an output signal in response to the exposure to the agent. The intensity of the signal is a function of the concentration of the agent. The titer or concentration of the agent is determined as the highest dilution that still elicits a measurable response in the output. The higher the number of dilutions tested, the higher the accuracy of the value obtained for the titer.

**Please amend the paragraph on page 59, lines 20-25, as follows:**

This plot represents the time necessary for the intensity of the output signal to reach the threshold value $\beta$ as a function of the concentration of the

biological agent used. [T his]This is a standard plot and will be used to determine the unknown concentration of the biological agent under study by measuring the time that a given dilution of it needs to give an output signal whose intensity equals the threshold $\beta$.

**Please amend the paragraph page 59, lines 31-32, as follows:**

The t$\beta$ value recorded above is entered into the standard plot obtained [above)]above and the corresponding ***c*** value is obtained.

**Please amend the paragraph on page 60, lines 3-33, as follows:**

**Example of the Real Time Virus [Titerng]Titering RTVT™ titering method**

Rat-2 cells were infected with serial dilutions of a reference preparation of a retroviral vector carrying the green fluorescent protein (GFP) gene (vector pSI-EGFPI see, Ropp *et al.* (1995) *Cytometry 21*:309-317). At increasing times after infection, the level of expression of the transgene was determined (as the level of fluorescence due to the GFP gene) as the output signal.

Table 3 represents the values obtained:

| Concentration (1 = $10^6$ particles/ml) | Time after infection (hrs) | Output signal fluorescence |
|---|---|---|
| 0.1 | 16 | 20.4 |
| | 24 | 30.1 |
| | 40 | 95.1 |
| | 48 | 138.7 |
| | 64 | 157.3 |
| 0.25 | 16 | 26.8 |
| | 24 | 48.5 |
| | 40 | 173.3 |
| | 48 | 228.2 |
| | 64 | 191.7 |
| 0.5 | 16 | 38.1 |
| | 24 | 72 |
| | 40 | 198.7 |
| | 48 | 296.2 |
| | 64 | 203.7 |

The threshold value of $\beta$ = 100 was arbitrarily selected for this example. The time ($t\beta$) necessary for the output signal to reach the threshold $\beta$ , for every concentration is shown in table 4.

**Please amend the paragraph on page 61, lines 7-21, as follows:**

As discussed above, TREE is a method for titering and standardization of preparations of viruses, vectors, antibodies, libraries, proteins, genes and any other moiety that is detectable based upon a output signal, such as fluorescence. The TREE method is an improvement of the Real Time Virus Titering (RTVT) method (see, International PCT application No. PCT/FR01/01366 published as International PCT application No. WO 01/186291). It is performed <u>with</u> a reporter moiety, such as a reporter virus (with a known titer) and the test sample (with unknown titer). The reporter, such as a reporter [virus]<u>virus,</u> has a readily detectable output signal that can be measured as a function of time. The effect of the moiety, such as a virus, of unknown titer is assessed. The moiety whose titer is assessed either increases or decreases the output signal as a function of time. This change in signal is used to assess the amount or concentration of the moiety of unknown concentration, and hence its titer.

**Please amend the paragraph beginning on page 63, line 26, through page 64, line 2, as follows:**

An arbitrary one value for FI (see Fig. 2B, 6 x $10^6$ FI units), typically, though not necessarily, near the greatest separation among the curves so that the numbers are readily discernable, was selected. The point at which each of the curves intersect this value is beta time ($t\beta$) for that combination of amounts of reporter plus dilution of the virus of unknown titer. $t\beta$, taken from the FI vs. Time (hrs) curves, for each sample containing a dilution of the unknown plus $10^6$ ip of the reporter virus is set forth in column [2]<u>3</u> of Table 2 below.

**Please amend the paragraph on page 66, lines 12-18, as follows:**

The infectious particles (ip) (infectious units, transducing units, etc.) are evidenced by the changes observed in the infected cells (vector DNA replication,

provirus integration, cell lysis, transgene expression and other observable [parameters]parameters). Infectious particles (ip) measures the number of particles effective in performing a process whose output is being measured; not all particles participate or are capable of participating in all processes.

**Please amend the paragraph beginning on page 69, line 7, through page 70, line 6, as follows:**

The method includes the steps of:

(a) preparation, for each biological agent, of a sample scale, obtained by a serial dilution of the biological agent at a R1 concentration,

(b) incubation of each sample of the dilution scale obtained in [1](a), with the target cells at a constant concentration R2,

(c) determination of the P product from the reaction R1 + R2, at a t moment, in each the sample; and

(d) realization of a theoretical curve H from the experimental points R1 and P, for each biological agent by iterative approximation of parameters of the reaction R1 + R2 → P, at the t moment, in accordance with this equation:

$$P = P_{max} (\pi R1)^r / (\kappa + (\pi R1)^r) \qquad r = 1,\ldots,n \qquad (2)$$

in which:

R1 represents the biological agent concentration in a sample from the scale;

R2 is concentration of target cells (*in vitro* or *in vivo*)

P (output) represents the product from the reaction R1 + R2 at a t moment;

$P_{max}$ represents the reaction maximal capacity;

$\kappa$ represents, at a constant R2 concentration, the resistance of the biological system for responding to the biological agent (resistance constant R2);

r represents a coefficient that depends on R1 and corresponds to the Hill coefficient; and

$\pi$ represents the intrinsic power of the R1 biological agent to induce a response in the biological system (P production at the t moment), and

(e) sorting the $\kappa$ and $\pi$ values obtained in (d) for each biological agent and the biological agent, and then ranking according to the values thereof.

**Please amend the paragraph beginning on page 74, line 18, through page 75, line 3, as follows:**

$\eta$ measures the internal heterogeneity of the reaction process under study. Complex processes include a huge chain of individual and [causally]causal events inside a multidimensional network of interrelated and interregulated biological reactions. Thus, the constant of resistance ($\kappa$) for the particular reaction process under study is a macroscopic indicator of the global resistance of that process ($\kappa = a1\kappa1 + a2\kappa2 + ...an\kappa n/n$). If the contribution of the individual microscopic constants of resistance ($a1\kappa1$, $a2\kappa2$, ...$an\kappa n$) for the individual steps involved in the process were homogeneous and no thresholds were present from one step to the next, then , no discontinuities in the increase of the Hill coefficient (i.e. in the change of $\kappa$) with R1 should be observed. The existence of a major heterogeneity among the $\kappa i$ values corresponding to the microscopic individual steps (i.e. the existence of thresholds for the intermediate steps) might lead to a macroscopic discontinuity in the system. Heterogeneity would cause a change in the rate of variation of the Hill coefficient and, which would require a jump in the macroscopic value of $\kappa$ in order for equation (2) to fit the data.

**Please amend the paragraphs beginning on page 79, line 10, through page 80, line 2, as follows:**

pNB-Adeno, which encodes the entire E2A and E4 regions and VA RNA I and II genes of Adenovirus type 5, was constructed by ligating into the polylinker of multiple cloning site of pBSII KS (+/-) (Stratagene, San Diego, USA) the SalI-HindIII fragment (9842-11555 nt) of Adenovirus [type 5)]type 5 and the BamHI-ClaI fragment (21563- 35950) of pBR325. All fragments of

adenovirus gene were obtained from the plasmid pBHG-10 (Microbix, Ontario, Canada). pNB-AAV encodes the genes rep and cap of AAV-2 and was [constructing]constructed by ligation of XbaI-XbaI PCR fragment containing the genome of AAV-2 from nucleotide 200 to 4480 into XbaI site of polylinker MCS of pBSIIKS(+/-). The PCR fragment was obtained from pAV1 (ATCC, USA). Plasmid pNB-AAV was derived from plasmid pVA1I, which contains the AAV genomic region, rep and cap. pNB-AAV does not contain the AAV ITR's present in pAV1. pAAV-CMV(nls)LacZ was provided by Dr Anna Salvetti (CHU, Nantes).

Plasmid pCMV(nls)LacZ (rAAV vector plasmid) and pNB-Adeno were prepared [on]in DH5a E.coli and purified by Nucleobond AX PC500 Kit (Macherey-Nagel), according to standard procedures. Plasmid pAAV-CMV(nls)LacZ is derived from plasmid psub201 by deleting the rep-cap region with SnaB I and replacing it with an expression cassette harboring the cytomegalovirus (CMV) immediate early promoter (407 bp), the nuclear localized $\beta$-galactosidase gene and the bovine growth hormone polyA signal (324 bp) (see, Chadeuf *et al.* (2000) *J. Gene Med. 2*:260-268. pAAV-CMV(nls)LacZ was provided by Dr Anna Salvetti.

**Please amend the paragraph on page 84, lines 3-20, as follows:**

To identify candidate amino acid (aa) positions on the rep protein involved in rep protein activity an Ala-scan was performed on the rep sequence. For this, each amino acid in the rep protein sequence was replaced with Alanine. To do this sets of rAAV that encode mutant rep proteins in which each differs from wild type by replacement of one amino acid with Ala, [was]were generated. Each set of rAAV was individually introduced into cells in a well of a microtiter plate, under conditions for expression of the rep protein. The amount of virus that could be produced from each variant was measured as described below. Briefly, activity of Rep was assessed by determining the amount of AAV or rAAV produced using infection assays on HeLa Rep-cap 32 cells and by measurement of AAV DNA replication using Real Time PCR, or by assessing

transgene ($\beta$-galactosidase) expression. The relative activity of each individual mutant compared to the native protein was assessed and "hits" identified. Hit positions are the positions in the mutant proteins that resulted in an alteration (selected to be at least about 20%), in this instance all resulted in a decrease, in the amount of virus produced compared to the activity of the native (wildtype) gene (see Fig. 3A).

**Please amend the paragraph beginning on page 88, line 8, through page 91, line 5, as follows:**

The hits in other AAV serotypes (see, also FIG. [4]5A and 5B) are as follows:

| HIT POSITION | | | | | | |
|---|---|---|---|---|---|---|
| AAV-2 | AAV-1 | AAV-3 | AAV-3B | AAV-4 | AAV-6 | AAV-5 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 22 | 22 | 22 | 22 | 22 | 22 | 22 |
| 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| 38 | 38 | 38 | 38 | 38 | 38 | 38 |
| 39 | 39 | 39 | 39 | 39 | 39 | 39 |
| 54 | 54 | 54 | 54 | 54 | 54 | 54 |
| 59 | 59 | 59 | 59 | 59 | 59 | 59 |
| 64 | 64 | 64 | 64 | 64 | 64 | 64 |
| 74 | 74 | 74 | 74 | 74 | 74 | |
| 86 | 86 | 86 | 86 | 86 | 86 | 85 |
| 88 | 88 | 88 | 88 | 88 | 88 | 87 |
| 101 | 101 | 101 | 101 | 101 | 101 | 100 |
| 124 | 124 | 124 | 124 | 124 | 124 | 123 |
| 125 | 125 | 125 | 125 | 125 | 125 | 124 |
| 127 | 127 | 127 | 127 | 127 | 127 | 126 |

| HIT POSITION | | | | | | |
|---|---|---|---|---|---|---|
| AAV-2 | AAV-1 | AAV-3 | AAV-3B | AAV-4 | AAV-6 | AAV-5 |
| 132 | 132 | 132 | 132 | 132 | 132 | 131 |
| 140 | 140 | 140 | 140 | 140 | 140 | |
| 161 | 161 | 161 | 161 | 161 | 161 | 158 |
| 163 | 163 | 163 | 163 | 163 | 163 | 160 |
| 175 | 175 | 175 | 175 | 175 | 175 | 172 |
| 193 | 193 | 193 | 193 | 193 | 193 | 190 |
| 196 | 196 | 196 | 196 | 196 | 196 | 193 |
| 197 | 197 | 197 | 197 | 197 | 197 | 194 |
| 221 | 221 | 221 | 221 | 221 | 221 | 217 |
| 228 | 228 | 228 | 228 | 228 | 228 | 224 |
| 231 | 231 | 231 | 231 | 231 | 231 | 227 |
| 234 | 234 | 234 | 234 | 234 | 234 | 230 |
| 237 | 237 | 237 | 237 | 237 | 237 | 233 |
| 250 | 250 | 250 | 250 | 250 | 250 | 246 |
| 258 | 258 | 258 | 258 | 258 | 258 | 254 |
| 260 | 260 | 260 | 260 | 260 | 260 | 256 |
| 263 | 263 | 263 | 263 | 263 | 263 | 259 |
| 264 | 264 | 264 | 264 | 264 | 264 | 260 |
| 334 | 334 | 334 | 334 | 334 | 334 | 330 |
| 335 | 335 | 335 | 335 | 335 | 335 | 331 |
| 337 | 337 | 337 | 337 | 337 | 337 | 333 |
| 341 | 341 | 341 | 341 | 341 | 341 | 337 |
| 342 | 342 | 342 | 342 | 342 | 342 | 338 |
| 347 | 347 | 347 | 347 | 347 | 347 | 342 |
| 350 | 350 | 350 | 350 | 350 | 350 | 346 |
| 354 | 354 | 354 | 354 | 354 | 354 | 350 |
| 363 | 363 | 363 | 363 | 363 | 363 | 359 |
| 364 | 364 | 364 | 364 | 364 | 364 | 360 |

| HIT POSITION | | | | | | |
|---|---|---|---|---|---|---|
| AAV-2 | AAV-1 | AAV-3 | AAV-3B | AAV-4 | AAV-6 | AAV-5 |
| 367 | 367 | 367 | 367 | 367 | 367 | 363 |
| 370 | 370 | 370 | 370 | 370 | 370 | 366 |
| 376 | 376 | 376 | 376 | 376 | 376 | 372 |
| 381 | 381 | 381 | 381 | 381 | 381 | 377 |
| 382 | 382 | 382 | 382 | 382 | 382 | 378 |
| 389 | 389 | 389 | 389 | 389 | 389 | 385 |
| 407 | 407 | 407 | 407 | 407 | 407 | 403 |
| 411 | 411 | 411 | 411 | 411 | 411 | 407 |
| 414 | 414 | 414 | 414 | 414 | 414 | 410 |
| 420 | 420 | 420 | 420 | 420 | 420 | 416 |
| 421 | 421 | 421 | 421 | 421 | 421 | 417 |
| 422 | 422 | 422 | 422 | 422 | 422 | 418 |
| 424 | 424 | 424 | 424 | 424 | 424 | 420 |
| 428 | 428 | 428 | 428 | 428 | 428 | 424 |
| 429 | 429 | 429 | 429 | 429 | 429 | 425 |
| 438 | 438 | 438 | 438 | 438 | 438 | 434 |
| 440 | 440 | 440 | 440 | 440 | 440 | 436 |
| 451 | 451 | 451 | 451 | 451 | 451 | 447 |
| 460 | 460 | 460 | 460 | 460 | 460 | 456 |
| 462 | 462 | 462 | 462 | 462 | 462 | 458 |
| 484 | 484 | 484 | 484 | 484 | 484 | 480 |
| 488 | 488 | 488 | 488 | 488 | 488 | 484 |
| 495 | 495 | 495 | 495 | 495 | 495 | 491 |
| 497 | 497 | 497 | 497 | 497 | 497 | 493 |
| 498 | 498 | 498 | 498 | 498 | 498 | 494 |
| 499 | 499 | 499 | 499 | 499 | 499 | 495 |
| 503 | 503 | 503 | 503 | 503 | 503 | 499 |
| 511 | 511 | 511 | 511 | 511 | 511 | 529 |

| HIT POSITION | | | | | | |
|---|---|---|---|---|---|---|
| AAV-2 | AAV-1 | AAV-3 | AAV-3B | AAV-4 | AAV-6 | AAV-5 |
| 512 | 512 | 512 | 512 | 512 | 512 | 530 |
| 516 | 516 | 516 | 516 | 516 | 516 | 534 |
| 517 | 517 | 517 | 517 | 517 | 517 | 535 |
| 518 | 518 | 518 | 518 | 518 | 518 | 536 |
| 519 | 519 | 519 | 519 | 519 | 519 | 537 |
| 542 | 543 | 542 | 542 | 542 | 543 | 561 |
| 548 | 549 | 548 | 548 | 548 | 549 | 567 |
| 598 | 599 | 600 | 600 | 599 | 599 | |
| 600 | 602 | 603 | 603 | 602 | 602 | 589 |
| 601 | 603 | 604 | 604 | 603 | 603 | 590 |

**Please amend the paragraph on page 91, lines 6-13, as follows:**

Sets of nucleic acids encoding the rep protein were generated. The rep proteins encoded by these sets of nucleic acid molecules were those in which each amino acid position [identied]identified as a "hit" in the ala-scan step, were each sequentially replaced by all remaining 18 amino acids using site directed mutagenesis. Each mutant was designed, generated, processed and analyzed physically separated from the others in addressable arrays. No mixtures, pools, nor combinatorial processing were used.

**Please amend the paragraph on page 92, lines 5-14, as follows:**

Also provided are combinations of the above mutant Rep 78, 68, [52.]52, 40 proteins, nucleic acids encoding the proteins, and recombinant AAV (any serotype) [contains]containing the mutation at the indicated position or corresponding position for serotypes other than AAV-2, including any set forth in the following table and corresponding SEQ ID Nos. Each amino acid sequence is set forth in a separate sequence ID listing; for each mutation or combination thereof there is a single SEQ ID setting forth the unspliced nucleic

acid sequence for Rep78/68, which for all mutations from amino acid 228 on, includes the corresponding Rep 52 and Rep 40 encoding sequence as well.

**Please amend the paragraph on page 111, lines 19-27, as follows:**

Mutant adeno-associated virus (AAV) Rep proteins and viruses encoding such proteins that include mutations at one or more of residues 64, 74, 88, 175, 237, 250 and 429, where residue 1 corresponds to residue 1 of the Rep78 protein [encoding]encoded by nucleotides 321-323 of the AAV-2 genome, and where the amino acids are replaced as follows: L by A at position 64; P by A at position 74; Y by A at position 88; Y by A at position 175; T by A at position 237; T by A at position 250; D by A at position 429 are provided. Nucleic acid molecules encoding these viruses and the mutant proteins are also provided.

**Please amend the paragraph beginning on page 112, line 26, through page 113, line 9, as follows:**

Also provided are nucleic acid molecules and rAAV (any serotype) in which position 630 (or the corresponding position in another serotype; see Figs. 5 and the table above) has been changed. Changes at this position and the region around it lead to changes in the activity or in the quantities of the Rep or Cap proteins and/or the amount of AAV or rAAV produced in cells transduced with AAV encoding such mutants. Such mutations include tgc to gcg change (SEQ ID No. 721). Mutations at any position surrounding the codon position 630 that increase or decrease the Rep or Cap proteins quantities or activities are also provided. Methods using the rAAV (any serotype) that contain nucleic acid molecules with a mutation at position 630 or within 1, 2, 3 . . . .10 or more bases thereof for the intracellular expression rep proteins or the rep gene mutants covered by claims 10 to 13, for the production of AAV or rAAV (either *in vitro*, *in vivo* or *ex vivo*) are provided. *In vitro* methods include cell free systems, expression or replication and/or virus assembly.

**IN THE CLAIMS:**

**Please amend claims 9, 13, 14, 24, 28, 31 and 41 with the following amended claims:**

9. (Amended) The method of claim 1, further comprising:

(d) modifying the nucleic acid molecules that encode the hits, to produce a set of nucleic acid molecules that encode modified hits;

(e) introducing the each set of nucleic acids that encode the modified hits into cells; and

(f) individually screening the sets of cells that contain the nucleic acid molecules that encode the modified hits to identify one or more cells that encodes a protein that has activity that differs from the target protein and has properties that differ from the original hits, wherein each such protein is designated a lead.

13. (Amended) The method of claim 2, wherein the nucleic acid molecules in step (a) are produced by [a] systematically changing each codon in the target protein to a pre-selected codon.

14. (Amended) the method of claim 13, wherein the codon is selected from a codon encoding Ala (A), Ser [(s)](S), Pro (P), or Gly (G).

24. (Amended) The method of claim [2]7, wherein at step (b) the titer of the viral vectors in each set of cells is assessed.

28. (Amended) The method of claim [26]27, wherein at step (f) the titer of the viral vectors in each set of cells is determined.

31. (Amended) The method of claim 30, wherein the Hill analysis, comprises:

(a) preparing a sample of each nucleic acid molecule or a plasmid or vector that comprises each nucleic acid molecule (biological agent), wherein each sample is obtained by a serial dilution of the molecules or vector or plasmid at a concentration R1,

(b) incubating each sample of the dilution obtained in (a) with the host cells (target cells) at a constant concentration R2,

(c) determining a P product from the reaction R1 + R2, at a t moment, in each the sample; and

(d) preparing a theoretical curve H from the experimental points R1 and P, for each biological agent by iterative approximation of parameters of the reaction R1 + R2 → P, at the t moment, in accordance with the equation:

$$P = P_{max} (\pi R1)^r / (\kappa + (\pi R1)^r) \qquad r = 1,\ldots,n \qquad (2)$$

in which:

R1 represents the biological agent concentration in a sample from the scale;

R2 is concentration of target cells (*in vitro* or *in vivo*)

P (output) represents the product from the reaction R1 + R2 at a t moment;

$P_{max}$ represents the reaction maximal capacity;

$\kappa$ represents, at a constant R2 concentration, the biological system for responding to the biological agent (resistance constant R2);

r represents a [dependant]<u>dependent</u> coefficient of R1 and corresponds to the Hill coefficient; and

$\pi$ represents the intrinsic power of the R1 biological agent to induce a response in the biological system (P production at the t moment), and

(e) sorting the $\kappa$ and $\pi$ values obtained in (d) for each protein encoded by the nucleic acid molecules or plasmids or vectors and the cells, and then ranking according to the values thereof.

41. (Amended) A method for producing a protein having modified properties, comprising:

(a) preparing a population of nucleic acid molecules that encode rationally modified proteins;

(b) inserting the population into [of] expression vectors;

(c) introducing each vector into host cells therefor, and expressing the modified proteins,

(d) screening each modified protein, and selecting one or more that has (have) a modified property.